

Is Working Memory Training Effective? A Meta-Analytic Review

Monica Melby-Lervåg
University of Oslo

Charles Hulme
University College London and University of Oslo

It has been suggested that working memory training programs are effective both as treatments for attention-deficit/hyperactivity disorder (ADHD) and other cognitive disorders in children and as a tool to improve cognitive ability and scholastic attainment in typically developing children and adults. However, effects across studies appear to be variable, and a systematic meta-analytic review was undertaken. To be included in the review, studies had to be randomized controlled trials or quasi-experiments without randomization, have a treatment, and have either a treated group or an untreated control group. Twenty-three studies with 30 group comparisons met the criteria for inclusion. The studies included involved clinical samples and samples of typically developing children and adults. Meta-analyses indicated that the programs produced reliable short-term improvements in working memory skills. For verbal working memory, these near-transfer effects were not sustained at follow-up, whereas for visuospatial working memory, limited evidence suggested that such effects might be maintained. More importantly, there was no convincing evidence of the generalization of working memory training to other skills (nonverbal and verbal ability, inhibitory processes in attention, word decoding, and arithmetic). The authors conclude that memory training programs appear to produce short-term, specific training effects that do not generalize. Possible limitations of the review (including age differences in the samples and the variety of different clinical conditions included) are noted. However, current findings cast doubt on both the clinical relevance of working memory training programs and their utility as methods of enhancing cognitive functioning in typically developing children and healthy adults.

Keywords: working memory training, ADHD, attention, learning disabilities

Working memory is one of the most influential theoretical constructs in cognitive psychology. This influence derives, at least in part, from links between measures of working memory capacity and a wide variety of real world skills (e.g., Cohen & Conway, 2008), as well as applications to issues in cognitive development and developmental cognitive disorders (for a review, see Gathercole & Alloway, 2006). Recently, excitement has been generated by claims that working memory capacity can be trained (e.g., Diamond & Lee, 2011; Klingberg, 2010; Morrison & Chein, 2011). Such a result would have important theoretical and practical implications. To assess these claims, in this article, we present a systematic meta-analytic review of studies that have examined the effects of working memory training in both children and adults.

The Nature of Working Memory

Working memory has been defined as “a brain system that provides temporary storage and manipulation of the information

necessary for . . . complex cognitive tasks” (Baddeley, 1992, p. 556). Historically, the concept of working memory evolved from earlier concepts of short-term memory. Short-term memory was initially seen as a limited capacity memory store that was subject to rapid loss due to decay (Atkinson & Shiffrin, 1968). A number of studies have shown that measures of short-term memory, such as digit span, correlate modestly with measures of higher level cognitive function, such as IQ (Mukunda & Hall, 1992; Unsworth & Engle, 2007b), reading (Swanson, Zheng, & Jerman, 2009), and arithmetic skills (Swanson & Jerman, 2006). In contrast to short-term memory tasks such as digit span, working memory tasks involve trying to maintain information in active memory while simultaneously performing distracting or interfering activities (Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980). In this sense working memory capacity could be seen as a limit on an individual’s ability to repeatedly retrieve information from permanent or secondary memory that has been lost from the focus of attention due to competing cognitive activity (for a review, see Conway et al. 2005; Unsworth & Engle, 2007a, 2007b). Measures of working memory consistently show higher correlations with measures of higher level cognitive functions than do simple memory span tasks (for a review, see Engle, 2002).

In practice, a wide range of tasks involving both verbal and nonverbal materials have been used to assess working memory skills (see, for example, Alloway, Gathercole, & Pickering, 2006; Kane et al., 2004). It has been debated as to whether working memory capacity reflects separable, modality specific (verbal versus visual) systems or a domain-general cognitive capacity. Evidence from large-scale latent variable studies with both children (Alloway et al., 2006) and adults (Kane et al., 2004) supports the

This article was published Online First May 21, 2012.

Monica Melby-Lervåg, Department of Special Needs Education, University of Oslo, Oslo, Norway; Charles Hulme, Division of Psychology and Language Sciences, University College London, London, England, and Department of Special Needs Education, University of Oslo.

Correspondence concerning this article should be addressed to Monica Melby-Lervåg, Department of Special Needs Education, University of Oslo, Pb 1140 Blindern, 0318 Oslo, Norway, or to Charles Hulme, Division of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 2PF, England. E-mail: monica.melby-lervag@isp.uio.no or c.hulme@ucl.ac.uk

conclusion that working memory capacity is best thought of as predominantly a domain-general capacity (though specific working memory tasks may show small degrees of modality specificity in their storage demands). The capacity tapped by these multiple measures of working memory capacity is typically conceptualized as reflecting some general limitation on attentional capacity (Engle, 2002).

This conceptualization of working memory capacity as a general limitation on attentional capacity was perhaps first clearly articulated by Engle, Tuholski, Laughlin, and Conway (1999). These authors gave a number of working memory tasks, together with a number of conventional short-term memory tasks (digit span and word span) and tests of general fluid intelligence (gf), to a large group of adults. They showed that measures of working memory were separate from (though correlated with) measures of short-term memory. They argued that what was shared between the working memory and the short-term memory measures reflected memory storage, and the unique variance measured by working memory tasks was executive attention. When the variance common to working memory and short-term memory was statistically removed from the working memory measures, these measures still correlated well with gf, but when the variance common to working memory and short-term memory (STM) was removed from the STM measures, these measures no longer correlated with gf. These and similar findings led Engle (2002) to argue that the working memory construct is "related to, maybe isomorphic to, general fluid intelligence and executive attention" (p. 22).

It should be noted that this notion of working memory capacity as synonymous with executive attention in turn leads to the view that individuals' with high working memory capacity will perform better on tasks requiring the inhibition of distracting information. A number of lines of evidence support this idea. For example, individuals with high working memory capacity perform better on an "antisaccade task" in which they have to inhibit an eye movement towards a visual cue that occurs on the opposite side of the screen to a brief stimulus that requires a judgment (Kane, Bleckley, Conway, & Engle, 2001). Similarly, the Stroop task requires participants to name the ink color in which a word is printed and to ignore the color word presented (e.g., say "red" when the word *blue* is presented written in red ink). On such incongruent trials, there is a strong tendency to respond with the color word, not the ink color. Kane and Engle (2003) showed that people with high working memory capacity found it easier to inhibit prepotent responses to the color words, but only in the more difficult condition when such incongruent trials were relatively infrequent (and hence, when the task goal may have slipped from the focus of attention). This result has also been replicated in children (Marco-vitch, Boseovski, & Knapp, 2007). People with high-working memory capacity also appear to be better at inhibiting distracting information in a dichotic listening task (Conway, Cowan, & Bunting, 2001).

In summary, working memory capacity is often regarded as tapping a domain general attentional resource limitation, so that working memory limitations are often associated with failures to maintain task focus and to inhibit the processing of, and responses to, distracting information. It is certainly the case that alternative, overlapping conceptualizations of working memory abound (e.g., working memory and inhibition may be seen as interdependent processes (Engle & Kane, 2004) or that working memory may be

seen as reflecting limitations in the ability to limit the processing of irrelevant information (Hasher, Lustig, & Zacks, 2007). However, the conception of working memory capacity as a domain general attentional resource limitation is the one that seems dominant in the literature on the working memory training effects considered here.

The Putative Role of Working Memory in Cognitive Development

The idea that working memory tasks estimate the limits of an individual's attentional capacity has led researchers to hypothesize that such a limit might be expected to have critical implications for cognitive development (see Case, 1985; Pascual-Leone, 1970). Furthermore, a working memory deficit has been invoked as a potential explanation for a variety of developmental cognitive disorders. In relation to reading disorders, a large number of studies have shown that children with reading problems have deficits on traditional, verbal short-term memory tasks (Melby-Lervåg, Lyster, & Hulme, 2012). Such findings have led to the suggestion that the efficient operation of phonological codes in memory is necessary for various phonological processes that are involved in learning to read words (Baddeley, 1986; Gathercole & Baddeley, 1993). Moreover, Swanson (2006) argued that working memory deficits (that are not restricted to traditional short-term memory tasks) are fundamental problems in children with reading disabilities. He claims that "we believe that reading disabled students' executive system (and more specifically monitoring activities linked to their capacity for controlled and sustained attention in the face of interference or distraction) is impaired" (Swanson, 2006, p. 83). Swanson argued that these broad working memory deficits contribute to problems in learning to read by creating problems in maintaining task relevant information, in suppressing task irrelevant information, and in accessing information from long-term memory.

Similarly, Passolunghi (2006) claimed that working memory problems are a central deficit in children with a mathematics disorder and that working memory plays a crucial role both in calculation and in solving arithmetic word problems (Passolunghi, 2006; Passolunghi & Siegel, 2001). Deficits in executive functioning, including working memory, have also been proposed as playing an important role in accounting for the symptoms of attention-deficit/hyperactivity disorder (ADHD) such as impairments of behavioral regulation, task planning, and selective attention (Klingberg et al. 2005; Mezzacappa & Buckner, 2010). Working memory problems have also been suggested to represent a key component in explaining the cognitive difficulties seen in children with autism spectrum disorder (Kenworthy, Yerys, Anthony, & Wallace, 2008) and specific language impairment (Archibald & Gathercole, 2006). For instance, Archibald and Gathercole (2006) claimed that

it seems likely that the striking deficits of children with specific language impairment in these two key domains [i.e., verbal short-term memory and verbal working memory] of immediate memory . . . make a major contribution to the learning difficulties experienced by these children. (p. 154)

However, a deficit in working memory capacity (executive control) is a very general explanation that seems insufficient by

itself as an explanation for such a wide variety of seemingly disparate disorders (see Hulme & Snowling, 2009). It appears necessary to supplement such explanations either by postulating additional deficits in each of the different disorders or perhaps by postulating that different forms of working memory deficit might cause different forms of disorders (Hulme & Snowling, 2009). For example, Geary, Hoard, Nugent, and Bailey (2011) found that longitudinal differences in the growth of arithmetic skills between children with and without arithmetic problems were predicted by variations in a range of skills including number processing, retrieval of number basic facts from long-term memory, and in-class attention, in addition to working memory deficits.

Training Working Memory Capacity: Theoretical Issues

If, as generally assumed, working memory reflects a general attentional resource limitation, this predicts that training working memory, if successful, should show transfer effects to untrained tasks (Shipstead, Redick, & Engle, 2010) because such training should lead to an increase in a domain-general attentional capacity that is critical for performing many diverse tasks. More specifically, in this view, working memory training would be expected to show both near- and far-transfer effects (see Barnett & Ceci, 2002). Near-transfer effects are effects on tasks close to those trained (e.g., improvements on a visuospatial working memory task following training on a verbal working memory task), whereas far-transfer effects are effects on tasks quite different from those trained (e.g., improvements on IQ tests following training on working memory tasks). Also, in line with theories that see working memory deficits as a potential explanation for a variety of developmental cognitive disorders such as reading disorder, mathematics disorder, ADHD, and specific language impairment, increases in working memory capacity might be expected to ameliorate the learning difficulties seen in these diverse groups of children. Therefore, theoretically, if one is able to train a domain-general working memory capacity successfully, far-transfer effects should be expected to occur to diverse skills and tasks that children may be struggling with (e.g., word decoding, arithmetic, attentional control, behavioral inhibition, and language abilities). This notion of transfer effects (see Chein & Morrison, 2011; Holmes et al., 2010; Klingberg, 2010; Perrig, Hollenstein, & Oelhafen, 2009) also explains the potential practical importance of working memory training, since it should transfer to other “real world” tasks, such as performing an IQ test, or to improved attentional skills that might have general effects on cognitive development and school attainment.

Many claims have been made that working memory training has quite general effects, with perhaps the most striking claim being that it can result in increases on standardized measures of intelligence such as the Raven's Progressive Matrices test (Raven, Raven, & Court, 2003). As outlined earlier, such far-transfer effects should be expected if working memory performance reflects principally the effects of a general-purpose attentional system. For example, Carpenter, Just, and Shell (1990) described Raven's Progressive Matrices as “a classic test of analytic intelligence . . . the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge derived from either schooling or previous

experience” (p 404). Furthermore, one of the major determinants of performance on this test, according to a formal model of performance developed by Carpenter et al., is the ability “to dynamically manage a large set of problem solving goals in working memory” (p 404). Such a view clearly leads to the prediction that working memory training programs, if they are effective, should give rise to improvements on attentionally demanding tasks such as Raven's matrices. It therefore appears particularly critical to assess the extent to which working memory training programs are effective in increasing scores on such a test. Such transfer effects are also critical in relation to demonstrating practical or clinical benefits from working memory training. If working memory training programs only have effects on tasks that are very similar to those that have been trained, this would undermine much of their proposed theoretical and practical importance.

Working Memory Training Programs

In recent years, several commercial, computer-based, working memory training programs have been developed. The most well-known is CogMed (<http://www.cogmed.com/>) which is available in 30 countries and is widely used in schools and clinics. This program is based on eight different exercises involving both visuospatial and verbal working memory tasks, in which the difficulty level varies adaptively during training. Other commercially available working memory training programs include Jungle Memory (<http://www.junglememory.com/>), which is based on three different tasks, and Cognifit (<http://www.cognifit.com/>), which is based on auditory, visual, and cross-modal working memory tasks. This review includes studies that have used all three programs as well as other research based computerized working memory training methods.

Some strong claims have been made about the effectiveness of two of these commercial programs. The Jungle memory website claims that the program will benefit children with ADHD, dyslexia and language impairments, dyspraxia and sensory integration difficulties, and autism spectrum disorders, as well as children with poor grades. It is claimed that “Jungle Memory improved IQ, working memory, and grades . . . Jungle Memory is the only brain training program proven to improve grades immediately after use” (<http://junglememory.com>). Similarly, the CogMed website claims that “CogMed Working Memory Training is a solution for individuals who are held back by their working memory capacity. That means several large groups: children and adults with attention deficits or learning disorders” and that “When you improve working memory, you improve fluid IQ . . . you will be better able to pay attention, resist distractions, self-manage, and learn” (<http://www.cogmed.com>).

It may be worth noting that these working memory training programs all involve adaptive tasks in which participants are given many memory trials to perform that are at or slightly above their current capacity. However, these programs do not appear to rest on any detailed task analysis or theoretical account of the mechanisms by which such adaptive training regimes would be expected to improve working memory capacity. Rather, these programs seem to be based on what might be seen as a fairly naïve “physical-energetic” model such that repeatedly “loading” a limited cognitive resource will lead to it increasing in capacity, perhaps somewhat analogously to strengthening a muscle by repeated use.

Previous Reviews

Several recent narrative reviews have addressed the effects of working memory and cognitive training programs (Boot, Blakely, & Simons, 2011; E. Dahlin, Bäckman, Neely, & Nyberg, 2009; Diamond & Lee, 2011; Klingberg, 2010; Morrison & Chein, 2011; Perrig et al., 2009; Shipstead et al., 2010; Takeuchi, Taki, & Kawashima, 2010). The conclusions drawn from these narrative reviews are highly variable. Some of the reviews concluded that working memory training has very promising prospects. For example, Morrison and Chein (2011) concluded that “the results from individual studies encourage optimism regarding working memory training as a tool for general cognitive enhancement” (p. 46) and Klingberg (2010) concluded that “the observed training effects suggest that working memory training could be used as a remediating intervention for individuals for whom low working memory capacity is a limiting factor for academic performance or in everyday life” (p. 317). In contrast, Shipstead et al. (2010) were less optimistic and stated that “as of yet, the results are inconsistent and likely to be driven by inadequate controls and ineffective measurement of the cognitive variables of interest” (p. 245). This variation in the conclusions drawn from current narrative reviews probably reflects the fact that there are very large variations in results across studies in this field. Some studies show very large effects on far-transfer measures (e.g., Klingberg, Forssberg, & Westerberg, 2002), while others show no far-transfer effects at all (e.g., Holmes, Gathercole, & Dunning, 2009). There is also considerable variability in how the studies included in the narrative reviews are selected (e.g., in some cases studies without a control group are included), and this may also help to explain why the reviews reach such disparate conclusions.

To clarify the picture, we believe it is necessary to conduct a meta-analysis that can synthesize the size of effects obtained from working memory training programs on both near and far-transfer measures. A meta-analysis will also allow us to identify outliers and to examine variables that may help to explain the variability in outcomes between studies. Establishing clearly the nature and size of effects produced by working memory training programs is of considerable theoretical and practical importance. Theoretically, clarifying this issue has potentially important implications for our understanding of the mechanisms of learning and learning disorders. On a practical level, such knowledge is also relevant to debates about methods for ameliorating learning disorders in both children and adults. In relation to practical applications of working memory training, it is critical to establish, if immediate training effects are obtained, how durable they are when assessed on delayed follow-up tests.

The Current Review

Scope and Aims of the Review

Given the potential practical and theoretical importance of claims that working memory capacity can be trained, we decided to conduct a systematic meta-analytic review of existing studies. We assess the extent to which working memory training has near-transfer effects, that is, benefits on other working memory tasks similar to those trained. More critically, however, we also assess far-transfer effects to tasks that have not been trained

directly (e.g., does training on working memory tasks improve performance on measures of reading or IQ?). Transfer effects are also critical in relation to demonstrating practical or clinical benefits from working memory training. If working memory training programs only have effects on tasks that are very similar to those that have been trained, this would undermine much of their proposed theoretical and practical importance. Also, in prior studies there is a large variation in the groups on which training effects are tested (children, young adults, and older adults, both unselected and from clinical groups) and in how the training is implemented (training duration and type of program). Since there are only a relatively small number of studies of working memory training, we adopted broad inclusion criteria but aimed to examine how differences between studies in sample characteristics (e.g., age, clinical vs. unselected groups) and design features affected their results.

Methodological Issues in Studies of Working Memory Training

The narrative review by Shipstead et al. (2010) made it clear that many studies that have examined the effects of working memory training have not always applied adequate methodological criteria that would allow training effects to be unambiguously demonstrated (e.g., Holmes et al., 2010; Mezzacappa & Buckner, 2010). However, the methodological requirements for an adequate study to demonstrate training effects on working memory are straightforward.

1. Any adequate study should have a random assignment of participants to the different groups. Random assignment serves to ensure that preexisting differences between participants cannot explain differences in outcome between groups.

2. The performance of a trained group needs to be compared with that of one or more suitable control groups (see below). In the absence of a control group, improvements between pretest and posttest in a trained group may simply reflect maturational changes, practice effects, or regression to the mean in studies that select participants for having low scores. Preferably, each group in a study should be tested before and after training. Analyzing changes between pretest and posttest scores across groups increases the power to detect a training effect. (Although, with random assignment, posttest scores alone may be interpretable, such a design is problematic unless large group sizes and robust randomization procedures are employed, see Shadish, Cook, & Campbell, 2002).

3. Ideally, an alternative active training procedure, delivered in exactly the same way, should be compared with the working memory training procedure. This controls for apparently irrelevant aspects of the training that might nevertheless affect performance. In a review of educational research Clark and Sugrue (1991) estimated that such Hawthorne or expectancy effects account for up to 0.3 standard deviations improvement in many studies. Studies that only compare working memory training with an untreated control group therefore run the risk that positive results may simply reflect expectancy effects. Although negative results from such trials would suggest that training is not effective, the reasons for such null results may be hard to interpret.

In our review, the studies included had to use a design that allowed training effects to be tested (i.e., have a pretest–posttest design with a training group and a control group). However, we

included both randomized and nonrandomized studies and studies with treated and untreated control groups. We used these variables in a moderator analysis to see how variations in the methodology used in the studies affected their results.

Method

The meta-analysis was designed and reported in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (www.prisma-statement.org). PRISMA is an international group of researchers in health care who have

developed a consensus statement for the conduct and reporting of systematic reviews and meta-analyses.

Literature Search, Inclusion Criteria, and Coding

Details concerning the method of literature search and criteria for inclusion and exclusion of studies are shown in Figure 1. To be included, a study had to use a working memory intervention and include standardized tests of nonverbal ability, verbal ability, attention, decoding, or arithmetic. Measures that involve problem solving primarily without relying on language were coded as

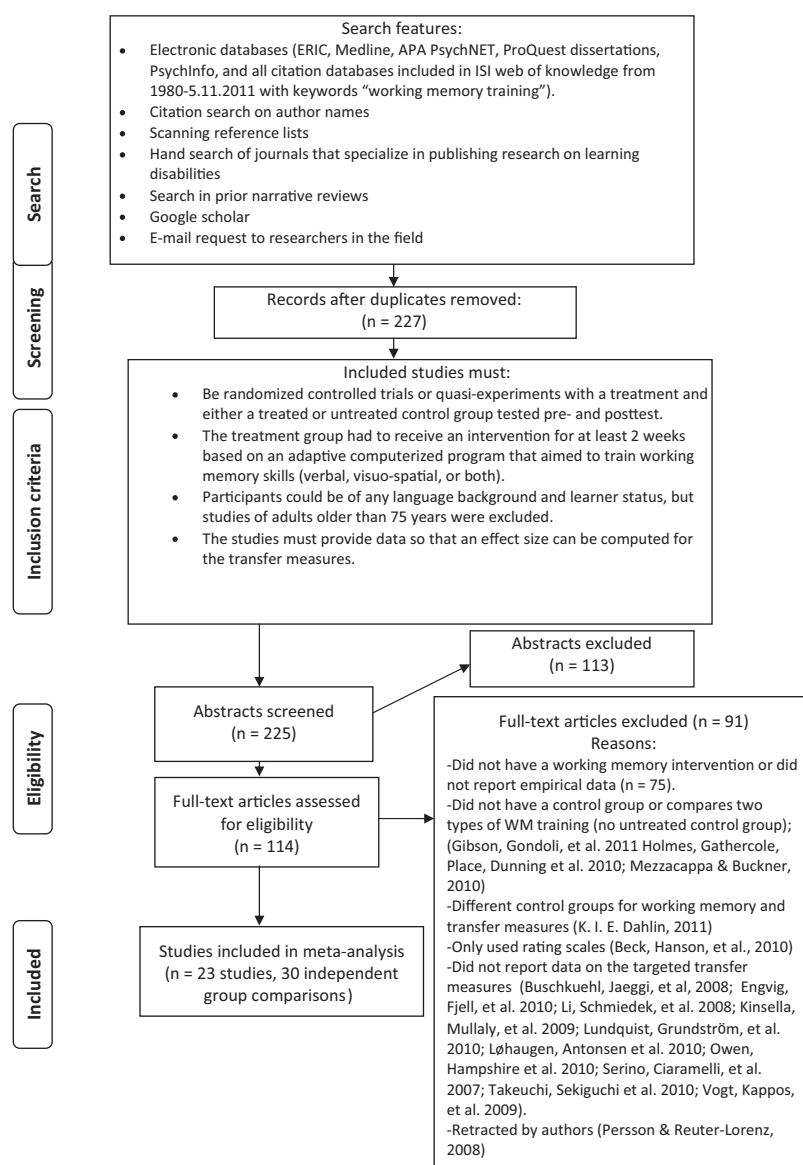


Figure 1. Flow diagram for the search and inclusion criteria for studies in this review. ISI = Institute for Scientific Information; WM = working memory; ERIC = Education Resources Information Center; APA = American Psychological Association. Adapted from "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," by D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and The PRISMA Group, 2009, *PLoS Med* 6(6). Copyright 2009 by the Public Library of Science.

nonverbal ability tests. Measures that involve comprehension and problem solving based primarily on verbal information were coded as measures of verbal ability. Measures that aimed to tap the participant's ability to concentrate selectively on one aspect of a task while ignoring others were coded as measures of attention. Notably, after going through the studies, it was clear that all studies that had measured processes related to attention had also included measures derived from the Stroop task (Stroop, 1935), and this task was therefore coded. The Stroop task is usually regarded as a measure of inhibitory processes in attention (see Smith & Jonides, 1999). We chose the Stroop task as our measure of attention in the meta-analysis simply because it was the one task that was included in all the studies included in the review.

The measures of reading (decoding) coded here included measures of the accuracy or fluency of word or nonword reading. Arithmetic measures included tests involving addition, subtraction, multiplication, and/or division. In addition, performance on working memory tests was also coded. Measures in which the participant was instructed to do a cognitive or motor task while concurrently remembering visual or spatial material were coded as measures of visuospatial working memory. Measures in which the participant was told to do a cognitive or motor task while concurrently remembering verbal material were coded as verbal working memory measures. In all cases in which there was more than one test of a construct, the average of the means and standard deviations for the tests were coded.

A sample of 50% of the studies was coded by two independent raters. The interrater correlation (Pearson's) for main outcomes was $r = .97$, 95% CI [.93, 1.00], $p < .0001$, and the agreement rate = 87.65%; the intercoder correlation for age was $r = .99$, 95% CI [.99, 1.00], $p < .0001$, and the agreement rate was 97.6%; and Cohen's kappa for categorical moderator variables was $\kappa = 0.88$, 95% CI [.75, .97], $p < .0001$, and the agreement rate = 97%. Any disagreements between raters were resolved by consulting the original article or by discussion.

Meta-Analytic Procedure

The analyses were conducted using the Comprehensive Meta-Analysis program (Borenstein, Hedges, Higgins, & Rothstein, 2005). Effect sizes were computed using Cohen's d , with corrections for small sample sizes (Hedges & Olkin, 1985). When Cohen's d is positive, the group receiving working memory training has the highest score. Cohen's d was calculated as the difference in gain (measured between pretest and posttest and at posttest immediately after training) between the training group and the control group and (when reported) for group differences in gain between the pretest and the follow-up test.

Overall effect sizes were estimated by calculating a weighted average of individual effect sizes using a random effects model. A 95% confidence interval was calculated for each effect size, to establish whether it was statistically significantly larger than zero.

Forest plots were used to examine the distributions of effect sizes and to detect outliers. A sensitivity analysis that allows an adjusted overall effect size to be estimated after removing studies one by one was undertaken to estimate the impact of outliers.

To examine the variation in effect sizes between studies, the Q -test of homogeneity was used (Hedges & Olkin, 1985). I^2 was also used in order to determine the degree of heterogeneity. I^2

assesses the percentage of between-study variance that is attributable to true heterogeneity rather than random error.

Funnel plots for random effects models were used to determine the presence of publication bias. In a funnel plot, a sample size dependent statistic is plotted on the y -axis and the effect size is plotted on the x -axis. In the absence of publication bias, this plot should form an inverted symmetrical funnel. Notably, when using a random effects model, the funnel plot can be difficult to interpret visually (Lau, Ioannidis, Terring, Schmid, & Olkin, 2006). Therefore, a trim and fill analysis was used in addition to the funnel plot. In the "trim and fill" method for random effects models (Duval & Tweedie, 2000), the impact of publication bias is estimated, and in the presence of publication bias, a trim and fill analysis can impute values in the funnel plot to make it symmetrical and calculate an adjusted overall effect size. Notably, when there are few studies, these procedures for analyzing publication bias becomes less reliable (see Cooper, Hedges, & Valentine 2009).

When we coded articles, it became clear that there were numerous instances of missing data. If data were critical to calculate an effect size, articles with missing data were excluded if authors did not respond to an e-mail request to provide the data (see inclusion criteria in flow chart). In cases in which an effect size could be computed on one outcome but data were missing on other outcomes or moderator variables, the study was included in all the analyses for which sufficient data were provided.

Moderator Variables

The ability of moderator variables to explain the variability in effect sizes between the studies was examined. The following moderator variables were used:

Age. The average age of participants in each study was coded. In the moderator analysis, due to a nonnormal distribution, it was not possible to analyze age as a continuous variable. Studies were therefore separated into three groups based on sample age: studies of younger children (under the age of 10 years), older children (11–18 years) and young adults (younger than 50 years), and older adults (51 years or older).

Training dose. The duration of the training (total number of hours in training) was coded. For the analysis, due to a nonnormal distribution, training duration was divided into studies with a total training duration up to and included 8 hr and studies with a total training time of 9 hr or more.

Design type. The procedure for separating participants into training and control groups was coded (randomized or nonrandomized).

Type of control group. The amount of attention and computer practice the control group received compared with the training group was coded (treated or untreated control group).

Learner status. Characteristics concerning the sampling of participants in the study were coded (whether they were sampled from a group with learning disorders or were unselected).

Intervention type. Characteristics concerning the training and intervention programs were coded.

Results

Information about all the studies included in the review is shown in Tables A1 and A2 (see Appendix). Table A1 shows sample age,

participant characteristics, design characteristics, and information about the training program for each study. Table A2 shows the sample size, the constructs and indicators coded from each study, and the effect size (Cohen's d) for pretest–posttest and pretest–delayed follow-up group differences in gains. As can be seen in Table A1, there are large variations between the studies in terms of both sample and design characteristics. It is also apparent in Table A2 that there were very large differences in the results from different studies. The moderator analysis is reported for pretest–posttest group differences on verbal working memory (see Table 1), visuospatial working memory (see Table 1), nonverbal ability (see Table 2), and attention (Stroop, see Table 3). For the remaining transfer measures and long-term effects, there were too few studies to allow meaningful moderator analyses to be performed.

Effects of Working Memory Training on Verbal Working Memory

Immediate training effects. Figure 2 shows the 21 effect sizes comparing pretest–posttest gains between working memory training groups and control groups on verbal working memory measures (N training groups = 707, mean sample size = 33.67, N controls = 641, mean sample size = 30.52). The mean effect size was large ($d = 0.79$), 95% CI [0.50, 1.09], $p < .001$. The heterogeneity between studies was significant, $Q(20) = 118.49$, $p < .001$, $I^2 = 83.12\%$. A sensitivity analysis showed that after removing outliers, the overall effect size ranged from $d = 0.72$, 95% CI [0.44, 1.00] to $d = 0.84$, 95% CI [0.55, 1.13]. The funnel plot indicated no publication bias, and in a trim and fill analysis, no

studies were imputed. Moderators of immediate training effects on verbal working memory are shown in Table 1 (first column). Age was the only significant moderator variable. Pairwise comparisons show that this difference was between younger children and older children, $Q(1) = 17.74$, $p < .001$, with younger children showing significantly larger benefits from training than do older children.

In summary, working memory training produces large immediate gains on measures of verbal working memory. There is considerable variation in the size of training effects across studies, with larger gains being shown in studies of younger children (below age 10 years) than in studies of older children (11–18 years).

Long-term training effects. Figure 3 shows the six effect sizes comparing pretest–posttest gains on verbal working memory measures between the working memory trained and control groups (N training groups = 135, mean sample size = 22.5, N controls = 118, mean sample size = 19.7). The mean effect size was small to moderate and nonsignificant ($d = 0.31$), 95% CI [−0.19, 0.80], $p = .22$. The heterogeneity between studies was significant, $Q(5) = 17.79$, $p < .01$, $I^2 = 71.90\%$. The follow-up measure of verbal working memory was taken on average some 9 months after the posttest. A sensitivity analysis showed that after removing outliers, the overall effect size ranged from $d = 0.10$, 95% CI [−0.30, 0.50], to $d = 0.47$, 95% CI [0.00, 0.94]. The funnel plot indicated no publication bias, and hence, no studies were imputed in a trim and fill analysis.

In summary, overall, the training effects on verbal working memory measures were not maintained at follow-up (9 months after training). It is notable that in the study by E. Dahlin, Nyberg, Bäckman, and Neely (2008), long-term effects were significant in

Table 1
Analysis of Moderators of Immediate Effects on Verbal Working Memory and Visuospatial Working Memory

Moderator variable	Verbal working memory				Visuospatial working memory			
	Number of effect sizes (k)	Effect size (d)	Heterogeneity (I^2)	Test of difference (Q test)	Number of effect sizes (k)	Effect size (d)	Heterogeneity (I^2)	Test of difference (Q test)
Age								
Young children	4	1.41**	63.23*		6	0.46**	26.86	
Children	6	0.26*	7.47		5	0.45**	54.24	
Young adults	7	0.74**	82.76**		4	0.61*	80.32**	
Older adults	4	0.95**	87.31	.00**	3	0.69	80.79**	.92
Training dose								
Large	12	0.94**	79.33**		10	0.49**	38.21	
Small	9	0.62**	87.74**	.31	8	0.53**	73.87**	.85
Design								
Nonrandomized	10	0.82**	88.19**		11	0.38**	34.93	
Randomized	11	0.76**	74.76**	.85	7	0.70**	72.53**	.20
Type of control								
Treated	8	0.99**	83.93**		10	0.63**	61.32**	
Untreated	12	0.69**	83.83**	.38	8	0.36**	45.68	.17
Learner status								
Learning disabled	7	0.56*	83.36**		9	0.47**	47.45	
Unselected	14	0.91**	80.73**	.63	9	0.57**	69.35**	.26
Intervention program								
CogMed	4	1.18**	82.67**		8	0.86*	24.12	
Jungle Memory	3	0.45	60.51		3	0.32	69.27	
N-back training	3	0.79	86.70**		—	—	—	
Other	8	0.75**	85.92**	.57	6	0.28*	60.03*	
Cognifit	—	—	—	—	2	0.44	0.00	.04*

Note. Dashes indicate no data were reported.

* $p < .05$. ** $p < .01$.

Table 2
Analysis of Moderators of Immediate Effects on Nonverbal Abilities

Moderator variable	Number of effect sizes (<i>k</i>)	Effect size (<i>d</i>)	Heterogeneity (I^2)	Test of difference (<i>Q</i> test)
Age				
Young children	6	0.03	0	.20
Children	5	−0.05	69.19	
Young adults	6	0.37**	0	
Old adults	6	0.27	71.15	
Training dose				
Large	14	0.23*	41.97	.63
Small	8	0.15	55.35	
Design				
Nonrandomized	11	0.34**	38.0	.06
Randomized	11	0.04	38.41	
Type of control				
Treated	10	0.00	43.90	.01**
Untreated	12	0.38**	14.71	
Learner status				
Learning disabled	5	0.14	70.95**	.68
Unselected	17	0.25**	26.73	
Intervention program				
CogMed	8	0.13	46.64	.53
N-back training	5	0.34	0	
Other	9	0.18	61.78*	

* $p < .05$. ** $p < .01$.

the original study, but in this study, the verbal working memory task was very similar to the ones on which the participants had been trained.

Effects of Working Memory Training on Visuospatial Working Memory

Immediate training effects. Figure 4 shows the 18 effect sizes comparing pretest–posttest gains between working memory training and control groups on visuospatial working memory measures (N training groups = 610, mean sample size = 33.89, N

controls = 469, mean sample size = 26.05). The mean effect size was moderate ($d = 0.52$), 95% CI [0.32, 0.72], $p < .001$. The heterogeneity between studies was significant, $Q(17) = 41.37$, $p < .001$, $I^2 = 58.91\%$. A sensitivity analysis showed that after removing outliers, the overall effect size ranged from $d = 0.44$, 95% CI [0.26, 0.62], to $d = 0.55$, 95% CI [0.34, 0.76]. The funnel plot indicated no publication bias, and hence, no studies were imputed in the trim and fill analysis.

Moderators of immediate training effects on visuospatial working memory are shown in Table 1 (second column). The only

Table 3
Analysis of Moderators of Immediate Effects on Stroop

Moderator variable	Number of effect sizes (<i>k</i>)	Effect size (<i>d</i>)	Heterogeneity (I^2)	Test of difference (<i>Q</i> test)
Age				
Young children	3	0.35	0	.99
Children	3	0.34	47.28	
Adults	3	0.33	25.33	
Training dose				
Large	5	0.38	31.98	.71
Small	5	0.28*	0	
Design				
Nonrandomized	3	0.48*	26.43	.50
Randomized	7	0.29*	0	
Type of control				
Treated	5	0.30*	0	.84
Untreated	5	0.35*	31.62	
Learner status				
Learning disabled	5	0.26	28.14	.52
Unselected	5	0.41*	0	
Intervention program				
CogMed	5	0.35	19.61	.87
Other	5	0.31*	0	

* $p < .05$.

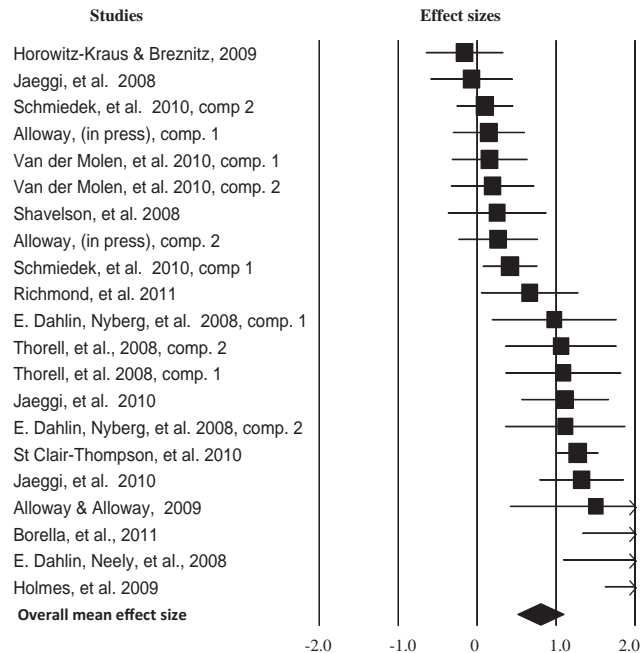


Figure 2. Forest plot for immediate training effects on verbal working memory, showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines; horizontal lines with arrows indicate that the confidence interval exceeds ± 2 Cohen's d). comp. = comparison.

significant moderator variable was program type. Pairwise comparisons showed that the CogMed training program demonstrated higher effect sizes than did the four noncommercial programs developed by researchers for the purposes of their studies. However, the difference between the commercially developed programs (CogMed, Cognifit, and Memory Booster) was not statistically significant, $Q(2) = 3.95$, $p = .14$.

In summary, working memory training produces moderately sized immediate gains on measures of visuospatial working memory, and there is little variation in effect sizes across studies. There

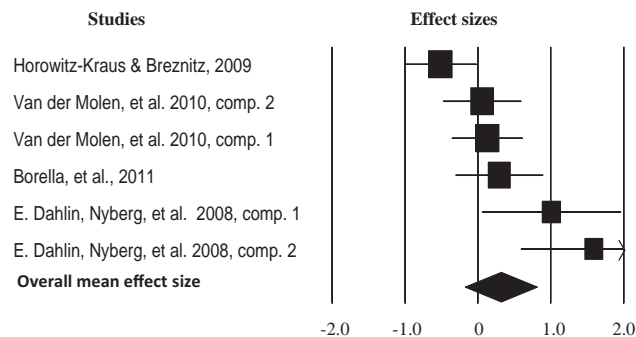


Figure 3. Forest plot for delayed training effects on verbal working memory showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines; horizontal lines with arrows indicate that the confidence interval exceeds ± 2 Cohen's d). comp. = comparison.

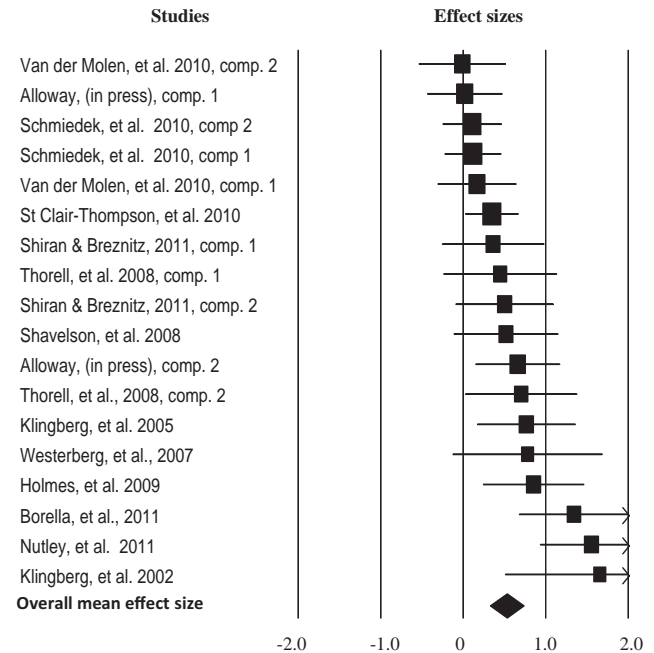


Figure 4. Forest plot for immediate training effects on visuospatial working memory showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines; horizontal lines with arrows indicate that the confidence interval exceeds ± 2 Cohen's d). comp. = comparison.

was also evidence that CogMed training produces higher effect sizes than the four programs developed by researchers for the purposes of their studies.

Long-term training effects. Figure 5 shows the four effect sizes comparing pretest–posttest gains between working memory training groups and control groups on visuospatial memory measures (N training groups = 102, mean sample size = 25.5, N controls = 94, mean sample size = 23.5). The mean effect size was moderate and significantly greater than zero ($d = 0.41$, 95% CI [0.13, 0.69], $p < .001$). The heterogeneity between studies was not significant, $Q(3) =$

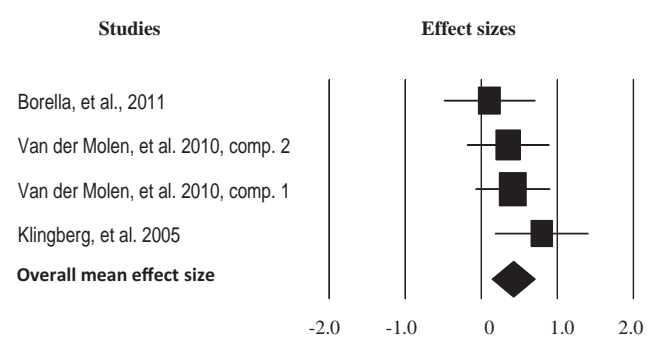


Figure 5. Forest plot for delayed training effects on visuospatial working memory showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines) for each study. comp. = comparison.

2.45, $p = .48$, $I^2 = 0\%$. On average, the follow-up measure was taken 5 months after the posttest (a shorter interval than for the verbal working memory). The funnel plot indicated no publication bias, and therefore, no studies were imputed.

In summary, superficially, the results here suggest that training effects on visuospatial working memory tasks are maintained at the delayed follow-up. However, in two out of the five studies assessed here (both from Van der Molen et al., 2010), the immediate training effects revealed were small (and in one case negative) effect sizes that subsequently increased to moderate effect sizes at follow-up. Arguably, such a pattern needs to be interpreted with caution (since it seems unlikely that genuine training effects would increase in size after the end of training). In the light of this, we would argue that we badly need further studies to assess whether visuospatial working memory tasks show genuine long-term benefits from working memory training.

Immediate Effects of Working Memory Training on Far-Transfer Measures

Nonverbal ability. Figure 6 shows the 22 effect sizes comparing the pretest–posttest gains between working memory training groups and control groups on nonverbal ability (N training groups = 628, mean sample size = 28.54, N controls = 528, mean sample size = 24.0). The mean effect size was small ($d = 0.19$), 95% CI [0.03, 0.37], $p = .02$. The heterogeneity between studies

was significant, $Q(21) = 39.17$, $p < .01$, $I^2 = 46.38\%$. The funnel plot indicated a publication bias to the right of the mean (i.e., studies with a higher effect size than the mean appeared to be missing), and in a trim and fill analysis, the adjusted effect size after imputation of five studies was $d = 0.34$, 95% CI [0.17, 0.52]. A sensitivity analysis showed that after removing outliers, the overall effect size ranged from $d = 0.16$, 95% CI [0.00, 0.32], to $d = 0.23$, 95% CI [0.06, 0.39].

Moderators of immediate transfer effects of working memory training to measures of nonverbal ability are shown in Table 2. There was a significant difference in outcome between studies with treated controls and studies with only untreated controls. In fact, the studies with treated control groups had a mean effect size close to zero (notably, the 95% confidence intervals for untreated controls were $d = -0.24$ to 0.22, and for treated controls $d = 0.23$ to 0.56). More specifically, several of the research groups demonstrated significant transfer effects to nonverbal ability when they used untreated control groups but did not replicate such effects when a treated control group was used (e.g., Jaeggi, Buschkuhl, Jonides, & Shah, 2011; Nutley, Söderqvist, Bryde, Thorell, Humphreys, & Klingberg, 2011). Similarly, the difference in outcome between randomized and nonrandomized studies was close to significance ($p = .06$), with the randomized studies giving a mean effect size that was close to zero. Notably, all the studies with untreated control groups are also nonrandomized; it is apparent from these analyses that the use of randomized designs with an alternative treatment control group are essential to give unambiguous evidence for training effects in this field.

In summary, we would emphasize that based on results from the most robust designs (randomized trials with an alternative treatment control group), there is no evidence of transfer effects from working memory training to measures of nonverbal ability.

Verbal ability. Figure 7 shows the eight effect sizes comparing pretest–posttest gains between working memory training groups and control groups on verbal ability (N training groups = 317, mean sample size = 39.63; N controls = 215, mean sample size = 26.87). The mean effect size was small and nonsignificant ($d = 0.13$), 95% CI [−0.09, 0.34]. There was no significant heterogeneity between studies, $Q(7) = 9.29$, $p = .23$, $I^2 = 24.64\%$; hence, the results were consistent across the studies in our sample. The funnel plot indicated no publication bias.

Stroop task (inhibitory processes in attention). Figure 8 shows the 10 effect sizes comparing the pretest–posttest gains between working memory training and control groups on the Stroop task (N training groups = 194, mean sample size = 19.4; N controls = 168, mean sample size = 16.8). The mean effect size was small to moderate ($d = 0.32$), 95% CI [0.11, 0.53], $p < .01$. There was no significant heterogeneity between studies, $Q(9) = 8.17$, $p = .51$, $I^2 = 0\%$; hence, the results were very consistent across the studies in our sample. There was no indication of publication bias.

Moderators of the effects of working memory training on the Stroop task are shown in Table 3. Note that due to the very small number of studies here, the power to detect significant differences between subsets of studies is low. Also, since $I^2 = 0\%$, there is little variation left to explain. This analysis should therefore be interpreted with caution.

Word decoding. Figure 9 shows the seven effect sizes comparing pretest–posttest gains between working memory training and

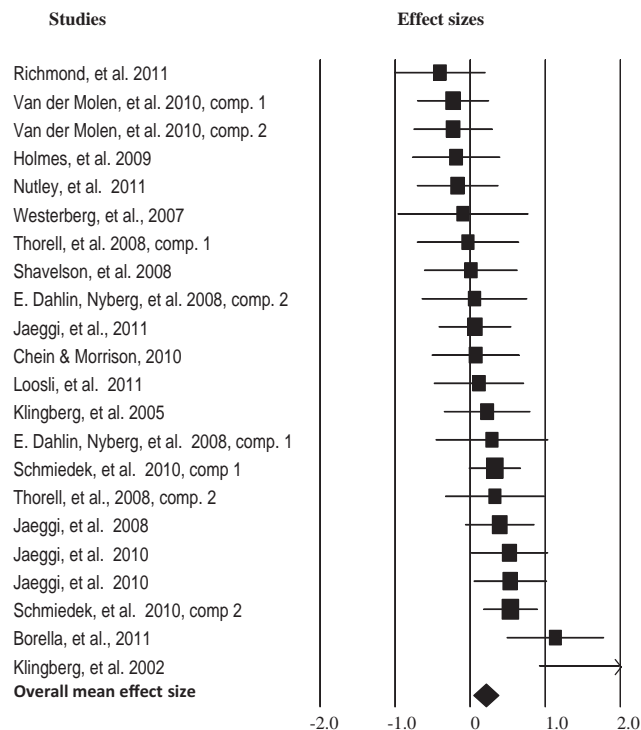


Figure 6. Forest plot for immediate training effects on nonverbal ability showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines; horizontal lines with arrows indicate that the confidence interval exceeds ± 2 Cohen's d). comp. = comparison.

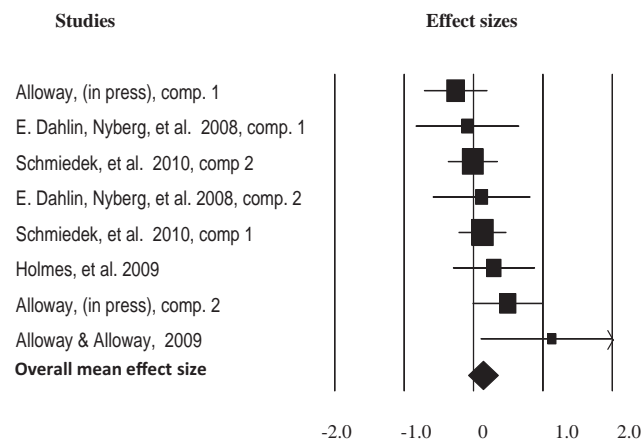


Figure 7. Forest plot for immediate training effects on verbal ability showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines; horizontal lines with arrows indicate that the confidence interval exceeds ± 2 Cohen's d).

control groups on word decoding (word and nonword reading; N training groups = 197, mean sample size = 28.14; N controls = 156, mean sample size = 22.28). The mean effect size was small and nonsignificant ($d = 0.13$), 95% CI $[-0.07, 0.35]$. There was no true heterogeneity between studies, $Q(6) = 2.65$, $p = .85$, $I^2 = 0\%$; hence, the results were consistent across the studies in our sample. The funnel plot indicated a bias to the left of the mean (studies with lower effect sizes than average). In a trim and fill analysis, one study was imputed, and the adjusted $d = 0.09$, 95% CI $[-0.10, 0.30]$.

Arithmetic. Figure 10 shows the seven independent effect sizes comparing pretest–posttest gains between working memory training groups and control groups on arithmetic (N training groups = 198, mean sample size = 28.28; N controls = 188, mean

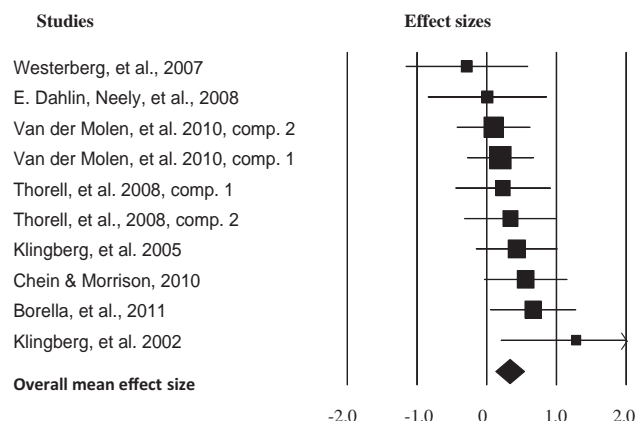


Figure 8. Forest plot for immediate training effects on the Stroop measure (inhibitory processes in attention) showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines; horizontal lines with arrows indicate that the confidence interval exceeds ± 2 Cohen's d). comp. = comparison.

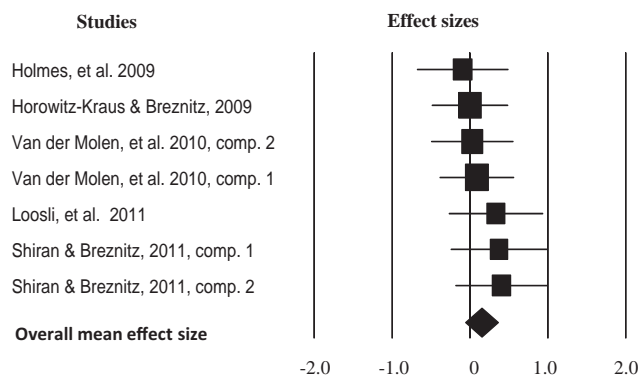


Figure 9. Forest plot for immediate training effects on word decoding showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines) for each study. comp. = comparison.

sample size = 26.86). The mean effect size was small and nonsignificant ($d = 0.07$), 95% CI $[-0.13, 0.27]$. There was no true heterogeneity between studies, $Q(6) = 2.80$, $p = .83$, $I^2 = 0\%$; hence, the results were consistent across the studies in our sample. There was no indication of publication bias.

Long-Term Effects of Working Memory Training on Transfer Measures

Table 4 shows the total number of participants in training and control groups, the total number of effect sizes, the time between the posttest and the follow-up, and the mean difference in gain between training and control groups from the pretest to the follow-up. It is apparent that all these long-term effects were small and nonsignificant. The true heterogeneity between studies was zero for all variables, indicating that the results were consistent across the studies included here. The funnel plot with trim and fill analyses did not indicate any publication bias. As for the attrition rate, on average, the studies lost 10% of the participants in the

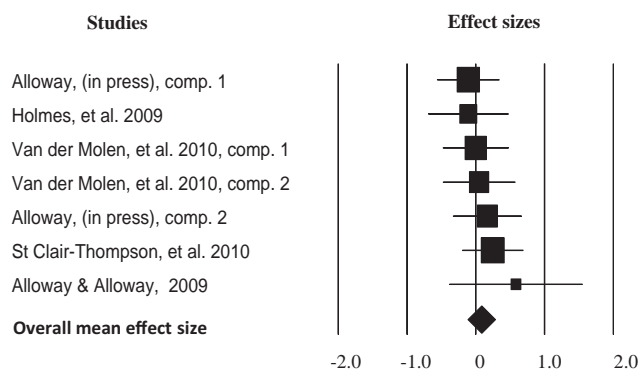


Figure 10. Forest plot for immediate training effects on arithmetic showing overall average effect size and confidence interval (Cohen's d , displayed as a diamond) and individual effect sizes (Cohen's d , displayed as a rectangle, with confidence intervals represented by horizontal lines) for each study. comp. = comparison.

Table 4

Total Number of Participants, Number of Effect Sizes, Time Between Posttest and Follow-Up, and Effect Size With 95% CI Between Pretest and Follow-Up

Variable	Pretest–follow-up group difference				
	Total N E (C)	Number of effect sizes (<i>k</i>)	Time between posttest and follow-up (months)	Effect size (<i>d</i>)	95% CI
Nonverbal ability	138 (120)	6	7.8	−0.06	−0.31, 0.17
Attention	102 (94)	4	5.0	0.09	−0.19, 0.37
Decoding	91 (84)	3	3.7	0.13	−0.17, 0.42
Arithmetic	108 (76)	3	3.33	0.18	−0.11, 0.47

Note. *N* = number of participants; *E* = experimental training group; *C* = control group; *CI* = confidence interval.

training group and 11% of the participants in the control group between the posttest and the follow-up. Only one study with two independent comparisons reported long-term effects for verbal ability (E. Dahlin et al., 2008). For the younger sample in this study, with 11 trained and seven control participants, long-term effects for verbal ability was nonsignificant ($d = 0.46$) 95% CI [−0.45, 1.37]. For the older participants in this study (13 trained, seven controls), the long term effects were negative and nonsignificant ($d = -0.08$), 95% CI [−0.96, 0.80].

In summary, there is no evidence from the studies reviewed here that working memory training produces reliable immediate or delayed improvements on measures of verbal ability, word reading, or arithmetic. For nonverbal reasoning, the mean effect across 22 studies was small but reliable immediately after training. However, these effects did not persist at the follow-up test, and in the best designed studies, using a random allocation of participants and treated controls, even the immediate effects of training were essentially zero. For attention (Stroop task), there was a small to moderate effect immediately after training, but the effect was reduced to zero at follow-up.

Discussion

Our meta-analysis of working memory training in children and adults reveals a clear pattern that has important theoretical and practical implications. Current training programs yield reliable, short-term improvements on both verbal and nonverbal working memory tasks. For verbal working memory, these short-term near-transfer effects are not sustained when they are reassessed after a delay averaging roughly 9 months. For visuospatial working memory, the pattern is less clear, and there is a suggestion that modest training effects may be present some 5 months after training, but the number of studies that this is based on is small. Most seriously, however, there is no evidence that working memory training produces generalized gains to the other skills that have been investigated (verbal ability, word decoding, or arithmetic), even when assessments take place immediately after training. For nonverbal reasoning, overall, there was a small but reliable improvement immediately after training. However, when we focus on studies using a robust design with treated controls and randomization, the effect size is zero. For attention (inhibition in the Stroop task), there is a small to moderate effect immediately after training, but the effect is reduced to zero at follow-up. Importantly, the pattern of results for transfer effects is highly consistent across

studies, with the heterogeneity between studies being virtually zero for all measures except verbal ability.

Methodological Issues in the Studies of Working Memory Training

When reviewing working memory training, it becomes clear that there are methodological shortcomings in many studies. Several studies were excluded because they lack a control group, since as outlined in the introduction, such studies cannot provide any convincing support for the effects of an intervention (e.g., Holmes et al., 2010; Mezzacappa & Buckner, 2010). However, among the studies that were included in our review, many used only untreated control groups. As demonstrated in our moderator analyses, such studies typically overestimated effects due to training, and research groups who demonstrated transfer effects when using an untreated control group typically failed to replicate such effects when using treated controls (Jaeggi, Buschkuhl, Jonides, & Shah, 2011; Nutley, Söderqvist, Bryde, Thorell, Humphreys, & Klingberg, 2011). Also, because the studies reviewed frequently use multiple significance tests on the same sample without correcting for this, it is likely that some group differences arose by chance (for example, if one conducts 20 significance tests on the same data set, the Type 1 error rate is 64% (Shadish, Cook, & Campbell, 2002). Especially if only a subset of the data is reported, this can be very misleading.

Finally, one methodological issue that is particularly worrying is that some studies show far-transfer effects (e.g., to Raven's matrices) in the absence of near-transfer effects to measures of working memory (e.g., Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Jaeggi et al., 2010). We would argue that such a pattern of results is essentially uninterpretable, since any far-transfer effects of working memory training theoretically must be caused by changes in working memory capacity. The absence of working memory training effects, coupled with reliable effects on far-transfer measures, raises concerns about whether such effects are artifacts of measures with poor reliability and/or Type 1 errors. Several of the studies are also potentially vulnerable to artifacts arising from regression to the mean, since they select groups on the basis of extreme scores but do not use random assignment (e.g., Holmes, Gathercole, & Dunning, 2009; Horowitz-Kraus & Breznitz, 2009; Klingberg, Forssberg, & Westerberg, 2002).

Practical Implications

Working memory training has sometimes been claimed to hold promise as a treatment for various forms of developmental disorders, including ADHD and reading disorders (Holmes, Gathercole, & Dunning, 2009; Horowitz-Kraus & Breznitz, 2009; Klingberg et al., 2005). The findings from the studies reviewed here are clear. Working memory training has positive effects on tasks close to those trained (near-transfer effects rather than far-transfer effects, see Barnett & Ceci, 2002). In all studies considered here, training has involved a variety of working memory tasks, and such training generalizes to other equivalent measures of working memory, but in no case is there evidence of a transfer to other less directly related tasks. This pattern of near-transfer effects in the absence of more general effects on cognitive performance (such as attention or nonverbal ability) or measures of scholastic attainment (reading or arithmetic ability) suggests that working memory training procedures cannot, based on the evidence to date, be recommended as suitable treatments for developmental disorders (such as ADHD or dyslexia).

It remains possible that training methods developed in the future will show better generalization, though current evidence is not encouraging in this regard. It also remains possible that these training programs, if applied to clinical groups of children (e.g., children with ADHD), would produce clear changes in specific symptoms, but so far, we lack evidence for this. We would note, however, that in some areas (for example in studies of children's reading and language difficulties) there is good evidence from randomized trials that "conventional" forms of treatment involving the direct training of reading and language skills are effective (e.g., Bowyer-Crane et al., 2008; Clarke, Snowling, Truelove, & Hulme, 2010). In the light of such evidence, it would seem very difficult to justify the use of working memory training programs in relation to the treatment of reading and language disorders. Finally, given that most of the studies included here involved samples of typically developing children and healthy adults, we would argue that our findings cast strong doubt on claims that working memory training is effective in improving cognitive ability and scholastic attainment in these groups.

Theoretical Implications

The pattern of near-transfer effects in the absence of far-transfer effects in studies of working memory training is hard to give a strong interpretation. This might be seen as a kind of null result, and it is impossible to rule out the possibility that future studies will be able to demonstrate far-transfer effects using different working memory training methods. However, it is equally possible that such far-transfer effects will not be forthcoming precisely because improvements in a modality independent working memory construct that is "related to, maybe isomorphic to, general fluid intelligence and executive attention" (Engle, 2002, p. 22) are difficult or impossible to produce with short-term programs rather restricted training programs. The training programs examined here are all of a relatively short duration (the mean duration of training across all studies was 12 hr). Furthermore, we would argue that current working memory training programs do not appear to be based on any clear theory of the processes involved or any clear task analysis. Rather, it seems these programs are based on what might be seen as a fairly naïve "physical-energetic" model: If you train a process (working memory), you will produce improvements in that process, perhaps by analogy with strengthening a muscle by exercising it.

The pattern of near-transfer effects in the absence of far-transfer effects documented in this review certainly needs to be interpreted with caution. Perhaps the most negative interpretation would be that the changes documented on near-transfer measures reflect very low-level changes in things like familiarity with specific tasks or even familiarity with being tested on a computer. A more interesting possible explanation, but one that current evidence does not necessitate, would invoke the idea that working memory training brings about some modality specific effects on memory processes per se. As outlined earlier, a guiding assumption has been that any conceivable working memory task should tap a "domain-free executive-attention system" as well as some memory-specific representational processes (perhaps the phonological and semantic representations of words (see, Melby-Lervåg & Hulme, 2010) that have to be remembered in a verbal working memory task. In this view, the pattern of near-transfer effects after working memory training documented here might be taken to suggest that these effects reflect modality specific (verbal or nonverbal) effects on memory processes rather than effects on a domain-free executive-attention system. We would emphasize, however, that this level of explanation is probably not necessary to deal with the evidence we have reviewed from current studies.

Also, it is worth noting that older studies have documented that intensive practice on certain memory tasks can lead to improvements on those tasks (e.g., Ericsson, Chase, & Faloon, 1980) that appear to be the product of task-specific strategies rather than a reflection of any generalizable improvements in "memory" per se. It remains a possibility that the near-transfer effects from working memory training demonstrated here reflect such task-specific strategy effects.

Finally, it might be argued that since the studies reviewed here typically only produce relatively short-lived task-specific improvements on working memory tasks, it is unreasonable to expect such training procedures to show transfer effects to other tasks such as reading and arithmetic. Our reason for evaluating such claims here is precisely because such transfer effects have often been claimed to be present in individual studies, and such effects have been claimed to be of practical importance by proponents of these programs. If reliable and durable effects of working memory training on working memory capacity can be demonstrated in future studies, a key question will be to assess possible transfer effects to other cognitive skills using psychometrically sound measures.

Limitations of the Current Meta-Analysis

Publication bias is potentially a serious threat to the validity of our meta-analysis of the effects of working memory training. Several methodological articles have shown that experimental studies of intervention effects such as those reviewed here are particularly vulnerable to publication bias. For example, Scherer, Langenberg, and von Elm (2007) traced roughly 3,000 randomized controlled trials presented at medical conferences and found that only approximately 60% ended up as published articles. The most important predictor of whether a study ended up being published was whether it demonstrated a positive result. Although we searched for gray literature (dissertations, conference proceeding, reports) and also contacted researchers in the field, we did not manage to retrieve any unpublished studies. Hence, in our meta-analysis, publication bias potentially represents a missing data problem. However, it is reasonable to suppose that this missing data problem would be likely to artificially

inflate our estimate of the effect of working memory training because it is to be expected that studies that fail to find an effect of working memory training are less likely to be published than studies showing positive effects of training. This was illustrated by Cuijpers, Smit, Bohlmeijer, Hollon, and Andersson (2010), who showed in a large set of therapy trials that the mean effect size was overestimated by $d = 0.25$ due to publication bias. It seems reasonable to suggest, therefore, that the true effects of current working memory training programs are likely to be even smaller than the effects estimated from our current meta-analysis.

A common, generic criticism of meta-analysis is that studies that are brought together differ in their characteristics and that when creating a summary of outcomes, important differences between studies may be ignored (see Bailer, 1997; Borenstein et al., 2009). However, it is important to note that one strength of meta-analysis is that the differences between studies can be addressed formally by examining the effects of moderator variables. In this review, in all cases in which there were a sufficient number of studies, we analyzed the impact of moderator variables on outcomes. As is apparent from these analyses (see Tables 3, 4, and 5), the results are very consistent across the different categories of moderators (the impact from moderator variables was significant in only 4 out of 25 different analyses). Also, for the far-transfer measures (both at posttest and follow-up test), it is crucial to note that the variation between the studies was essentially zero for measures of inhibitory processes in attention (Stroop), word decoding, and arithmetic and was small and nonsignificant for measures of verbal ability. In short, it is not likely that for any of the far-transfer measures there is appreciable systematic variation between the studies that can be explained by moderator variables. We should also note that the moderator analyses had to merge together participants from diverse clinical groups (including children with ADHD, dyslexia, and poor working memory and those with learning disabilities) simply because there were too few studies with clearly defined groups to make separating them meaningful. However, the very low degree of heterogeneity in the outcome on the far-transfer measures really does not encourage the view that there are large differences in outcome between these diverse groups.

Finally our meta-analyses combined studies of children and adults covering a wide range of age groups. Due to the small number of studies and the nonnormal distribution of age, in the moderator analyses using age, we had to split the sample of studies into four categories. For verbal working memory, younger children had reliably better training effects than did older children, while on the far-transfer measure, we found no reliable differences between the age categories in outcome. However, one might hypothesize that there could be developmental periods during which children are particularly sensitive to training (specifically, some might expect working memory training might be more effective in younger children, when neural systems are more plastic). In our moderator analysis, children under the age of 10 years are merged together in one category. This age category is very broad and might not be sensitive enough to capture whether the youngest children (say children below 7 years of age) show particularly strong effects of training. However, if one considers the two studies examining the youngest children (St. Clair-Thompson, Stevens, Hunt, & Bolder, 2010, children's mean age = 6 years 10 months; Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2009, children's mean age = 4 years 6 months), neither of these studies

demonstrates larger effect sizes on the far-transfer measures than the other studies that we considered (see Figures 6, 8, & 10).

Conclusions

Currently available working memory training programs have been investigated in a wide range of studies involving typically developing children, children with cognitive impairments (particularly ADHD), and healthy adults. Our meta-analyses show clearly that these training programs give only near-transfer effects, and there is no convincing evidence that even such near-transfer effects are durable. The absence of transfer to tasks that are unlike the training tasks shows that there is no evidence these programs are suitable as methods of treatment for children with developmental cognitive disorders or as ways of effecting general improvements in adults' or children's cognitive skills or scholastic attainments.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- *Alloway, T. P. (in press). Adaptive working memory training: Can it lead to gains in cognitive skills in students with learning needs? Can interactive working memory training improve learning? *Journal of Interactive Learning Research*.
- *Alloway, T. P., & Alloway, R. G. (2009). The efficacy of working memory training in improving crystallized intelligence. *Nature Precedings*. Retrieved from <http://precedings.nature.com/documents/3697/version/1/files/npre20093697-1.pdf>
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuo-spatial short-term and working memory in children: Are they separable? *Child Development*, 77, 1698–1716. doi:10.1111/j.1467-8624.2006.00968.x
- Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory in neurodevelopmental conditions* (pp. 139–160). Hove, England: Psychology Press.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 742–775). New York, NY: Academic Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Baddeley, A. (1992). Working memory. *Science*, 255, 556–559. doi: 10.1126/science.1736359
- Bailer, J. C. (1997). The promise and problems of meta-analysis. *New England Journal of Medicine*, 337, 559–561. doi:10.1056/NEJM199708213370810
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637. doi:10.1037/0033-2909.128.4.612
- Beck, S. J., Hanson, C. A., Puffenberger, S. S., Benninger, K. L., & Benninger, W. B. (2010). A controlled trial of working memory training for children and adolescents with ADHD. *Journal of Clinical Child and Adolescent Psychology*, 39, 825–836. doi:10.1080/15374416.2010.517162
- Boot, W., Blakely, D., & Simons, D. (2011). Do action video games improve perception and cognition? *Frontiers in Psychology*, 2(226), 1–6. doi:10.3389/fpsyg.2011.00226
- Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging*, 25, 767–778. doi:10.1037/a0020683
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (Version 2) [Software]. Englewood, NJ: Biostat.

- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley. doi:10.1002/9780470743386
- Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry*, 49, 422–432. doi:10.1111/j.1469-7610.2007.01849.x
- Buschkuhl, M., Jaeggi, S. M., Hutchison, S., Perrig-Chiello, P., Däpp, C., Müller, M., . . . Perrig, W. J. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging*, 23, 743–753. doi:10.1037/a0014342
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven's Progressive Matrices test. *Psychological Review*, 97, 404–431. doi:10.1037/0033-295X.97.3.404
- Case, R. (1985). *Intellectual development. Birth to adulthood*. New York, NY: Academic Press.
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386–404. doi:10.1016/0022-0965(82)90054-6
- *Chein, J. M., & Morrison, A. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17, 193–199. doi:10.3758/PBR.17.2.193
- Clark, R. E., & Sugrue, B. M. (1991). Research on instructional media, 1978–1988. In G. J. Anglin (Ed.), *Instructional technology: Past, present, and future* (pp. 327–343). Englewood, CO: Libraries Unlimited.
- Clarke, P. J., Snowling, M., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21, 1106–1116. doi:10.1177/0956797610375449
- Cohen, G., & Conway, M. A. (2008). *Memory in the real world* (3rd ed.). Hove, England: Psychology Press.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8, 331–335. doi:10.3758/BF03196169
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. doi:10.3758/BF03196772
- Cooper, H. M., Hedges, L. V., & Valentine, J. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.
- Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: Meta-analytic study of publication bias. *British Journal of Psychiatry*, 196, 173–178. doi:10.1192/bjp.bp.109.066001
- Dahlin, E., Bäckman, L., Neely, A. S., & Nyberg, L. (2009). Training of the executive component of working memory: Subcortical areas mediate transfer effects. *Restorative Neurology and Neuroscience*, 27, 405–419. doi:10.3233/rnn-2009-0492
- *Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, 23, 720–730. doi:10.1037/a0014296
- *Dahlin, E., Neely, A., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, 320, 1510–1512. doi:10.1126/science.1155466
- Dahlin, K. I. E. (2011). Effects of working memory training on reading in children with special needs. *Reading and Writing*, 24, 479–491. doi:10.1007/s11145-010-9238-y
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, 6, 122–149. doi:10.1016/S0022-5371(80)90312-6
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science*, 333, 959–964. doi:10.1126/science.1204529
- Duval, S., & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. doi:10.1111/j.0006-341X.2000.00455.x
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23. doi:10.1111/1467-8721.00160
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, 145–199). New York, NY: Elsevier.
- Engle, R. W., Tuholski, S. W., Laughlin, J., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable model approach. *Journal of Experimental Psychology: General*, 128, 309–331. doi:10.1037/0096-3445.128.3.309
- Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A., & Walhovd, K. B. (2010). Effects of memory training on cortical thickness in the elderly. *NeuroImage*, 52, 1667–1676. doi:10.1016/j.neuroimage.2010.05.041
- Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, 208, 1181–1182. doi:10.1126/science.7375930
- Gathercole, S. E., & Alloway, T. P. (2006). Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *Journal of Child Psychology and Psychiatry*, 47, 4–15. doi:10.1111/j.1469-7610.2005.01446.x
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2011). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*. Advance online publication. doi:10.1037/a0025398
- Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., Dobrzanski, B. A., Morrissey, R. A., & Thompson, A. N. (2011). Component analysis of verbal versus spatial working memory training in adolescents with ADHD: A randomized, controlled trial. *Child Neuropsychology* (October issue), 1–18. doi:10.1080/09297049.2010.551186
- Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- *Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, 12, F9–F15. doi:10.1111/j.1467-7687.2009.00848.x
- Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., & Elliott, J. G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, 24, 827–836. doi:10.1002/acp.1589
- *Horowitz-Kraus, T., & Breznitz, Z. (2009). Can error detection activity increase in dyslexic readers' brain following reading acceleration training? An ERP study. *PLoS ONE*, 4(9), e7141. doi:10.371/journal.pone.0007141
- Hulme, C., & Snowling, M. (2009). *Developmental cognitive disorders*. Oxford, England: Blackwell/Wiley.
- *Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6829–6833. doi:10.1073/pnas.0801268105
- *Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 10081–10086. doi:10.1073/pnas.1103228108
- *Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y.-F., Jonides, J., &

- Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence*, 38, 625–635. doi:10.1016/j.intell.2010.09.001
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity: Individual differences in memory span and the control of visual orienting. *Journal of Experimental Psychology: General*, 130, 169–183. doi:10.1037/0096-3445.130.2.169
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70. doi:10.1037/0096-3445.132.1.47
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189–217. doi:10.1037/0096-3445.133.2.189
- Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychological Review*, 18, 320–338. doi:10.1007/s11065-008-9077-7
- Kinsella, G. J., Mullaly, E., Rand, E., Ong, B., Burton, C., Price, S., . . . Storey, E. (2009). Early intervention for mild cognitive impairment: A randomised controlled trial. *Journal of Neurology, Neurosurgery & Psychiatry*, 80, 730–736. doi:10.1136/jnnp.2008.148346
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14, 317–324. doi:10.1016/j.tics.2010.05.002
- *Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., . . . Westerberg, H. (2005). Computerized training of working memory in children with ADHD: A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44, 177–186. doi:10.1097/00004583-200502000-00010
- *Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24, 781–791. doi:10.1076/jcen.24.6.781.8395
- Lau, J., Ioannidis, J. P. A., Terring, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal*, 333, 597–600. doi:10.1136/bmj.333.7568.597
- Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging*, 23, 731–742. doi:10.1037/a0014343
- Løhaugen, G. C. C., Antonsen, I., Håberg, A., Gramstad, A., Vik, T., Brubakk, A.-M., & Skranes, J. (2011). Computerized working memory training improves function in adolescents born at extremely low birth weight. *The Journal of Pediatrics*, 158, 555–561. doi:10.1016/j.jpeds.2010.09.060
- *Loosli, S., Buschkuhl, M., Perrig, W., & Jaeggi, S. (2011). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*. Advance online publication. doi:10.1080/09297049.2011.575772
- Lundqvist, A., Grundström, K., Samuelsson, K., & Rönneberg, J. (2010). Computerized training of working memory in a group of patients suffering from acquired brain injury. *Brain Injury*, 24, 1173–1183. doi:10.3109/02699052.2010.498007
- Marcovitch, S., Boseovski, J. J., & Knapp, R. J. (2007). Use it or lose it: Examining preschoolers' difficulty in maintaining and executing a goal. *Developmental Science*, 10, 559–564. doi:10.1111/j.1467-7687.2007.00611.x
- Melby-Lervåg, M., & Hulme, C. (2010). Serial and free recall in children can be improved by training: Evidence for the importance of phonological and semantic representations in immediate memory tasks. *Psychological Science*, 21, 1694–1700. doi:10.1177/0956797610385355
- Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138, 322–352. doi:10.1037/a0026744
- Mezzacappa, E., & Buckner, J. C. (2010). Working memory training for children with attention problems or hyperactivity: A school-based pilot study. *School Mental Health*, 2, 202–208. doi:10.1007/s12310-010-9030-9
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6(6), e1000097. doi:10.1371/journal.pmed1000097
- Morrison, A., & Chein, J. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, 18, 46–60. doi:10.3758/s13423-010-0034-0
- Mukunda, K. V., & Hall, V. C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence*, 16, 81–97. doi:10.1016/0160-2896(92)90026-N
- Nutley, S. B., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, 14, 591–601. doi:10.1111/j.1467-7687.2010.01022.x
- Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., . . . Ballard, C. G. (2010). Putting brain training to the test. *Nature*, 465, 775–778. doi:10.1038/nature09042
- Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, 32, 301–345. doi:10.1016/0001-6918(70)90108-3
- Passolunghi, M. C. (2006). Working memory and mathematical disability. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental condition* (pp. 113–138). Hove, England: Psychology Press.
- Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology*, 80, 44–57. doi:10.1006/jecp.2000.2626
- Perrig, W. J., Hollenstein, M., & Oelhafen, S. (2009). Can we improve fluid intelligence with training on working memory in persons with intellectual disabilities? *Journal of Cognitive Education and Psychology*, 8, 148–164. <http://dx.doi.org/10.1891/1945-8959.8.2.148>; <http://dx.doi.org/10.1891/1945-8959.8.2.148>
- Persson, J., & Reuter-Lorenz, P. A. (2008). Gaining control training executive function and far transfer of the ability to resolve interference. *Psychological Science*, 19, 881–888. doi:10.1111/j.1467-9280.2008.02172.x
- Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General overview*. San Antonio, TX: Harcourt Assessment.
- *Richmond, L. L., Morrison, A., Chein, J., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, 26, 813–822. doi:10.1037/a0023631
- Scherer, R., Langenberg, P., & von Elm, E. (2007). Full publication of results initially presented in abstracts. *Cochrane Database System Review*.
- *Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, 2, 1–10. doi:10.3389/fnagi.2010.00027
- Serino, A., Ciarra, E., Di Santantonio, A., Malagu, S., Servadei, F., & Ladavas, E. (2007). A pilot study for rehabilitation of central executive deficits after traumatic brain injury. *Brain Injury*, 21, 11–19. doi:10.1080/02699050601151811
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- *Shavelson, R. J., Yuan, K., Alonzo, A. C., Klingberg, T., & Andersson, M. (2008). On the impact of computerized cognitive training on working memory and fluid intelligence. In D. C. Berliner & H. Kuppermintz (Eds.), *Contributions of educational psychology to changing institutions, environments, and people* (pp. 1–11). New York, NY: Routledge.

- Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50, 245–276.
- *Shiran, A., & Breznitz, Z. (2011). The effect of cognitive training on recall range and speed of information processing in the working memory of dyslexic and skilled readers. *Journal of Neurolinguistics*, 24, 524–537. doi:10.1016/j.jneuroling.2010.12.001
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283, 1657–1661. doi:10.1126/science.283.5408.1657
- *St. Clair-Thompson, H. L., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, 30, 203–219. doi:10.1080/01443410903509259
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. doi:10.1037/h0054651
- Swanson, H. L. (2006). Working memory and reading disabilities: Both phonological and executive processing deficits are important. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (pp. 59–88). Hove, England: Psychology Press.
- Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, 76, 249–274. doi:10.3102/00346543076002249
- Swanson, H. L., Zheng, X. H., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities*, 42, 260–287. doi:10.1177/0022219409331958
- Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., . . . Kawashima, R. (2010). Training of working memory impacts structural connectivity. *Journal of Neuroscience*, 30, 3297–3303. doi:10.1523/JNEUROSCI.4611-09.2010
- Takeuchi, H., Taki, Y., & Kawashima, R. (2010). Effects of working memory training on cognitive functions and neural systems. *Reviews in the Neurosciences*, 21, 427–449. doi:10.1515/REVNEURO.2010.21.6.427
- *Thorell, L. B., Lindqvist, S., Bergman, S., Bohlin, N. G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, 12, 106–113. doi:10.1111/j.1467-7687.2008.00745.x
- Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132. doi:10.1037/0033-295X.114.1.104
- Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex spans and their relation to higher-order abilities. *Psychological Bulletin*, 133, 1038–1066. doi:10.1037/0033-2909.133.6.1038
- *Van der Molen, M. J., Van Luit, J. E., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research*, 54, 433–447. doi:10.1111/j.1365-2788.2010.01285.x
- Vogt, A., Kappos, L., Calabrese, P., Stöcklin, M., Gschwind, L., Opwis, K., & Penner, I. K. (2009). Working memory training in patients with multiple sclerosis—Comparison of two different training schedules. *Restorative Neurology and Neuroscience*, 27, 225–235. doi:10.3233/RNN-2009-0473
- *Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Ostensson, M. L., Bartfai, A., & Klingberg, T. (2007). Computerized working memory training after stroke—A pilot study. *Brain Injury*, 21, 21–29. doi:10.1080/02699050601148726

Appendix

Table A1

Characteristics of Studies of Working Memory Training Included in the Meta-Analysis

Study author & year	Age (years) E (C)	Design	Participant characteristics	Control treatment	Program type	Training duration
Alloway & Alloway (2009)	12.9 (13.0)	Pre-post randomized	Learning difficulties	Learning support with a special support staff (not computerized)	Jungle Memory	3 times per week, 30 min per session for 8 weeks
Alloway (in press)						
Comparison 1	10.6 (10.11)	Pre-post randomized	Learning difficulties	Practice as usual	Jungle Memory	Once per week for 8 weeks
Comparison 2	11.2 (10.11)	Pre-post randomized	Learning difficulties	Practice as usual	Jungle Memory	4 times per week for 8 weeks
Borella et al. (2011)	69.0 (69.15)	Pre-post randomized with follow-up	Healthy older adults	Treated (answering an autobiographic questionnaire)	Program developed for study, verbal working memory tasks	5 sessions completed within a 2-week time frame
Chein & Morrison (2010)	20.1 (20.6)	Pre-post randomized	Unselected	Untreated	Program developed for study, a combination of verbal and spatial WM tasks	30–45 min, 5 days per week over 4 weeks
E. Dahlin, Nyberg, et al. (2008)						
Comparison 1	23.67 (24.09)	Pre-post and follow-up (18 months after posttest), randomized	Unselected	Untreated	Study-developed program training WM of numbers, letters, colors, and spatial locations	15–45-min sessions over a period of 5 weeks
Comparison 2	68.4 (68.3)	Pre-post and follow-up (18 months after posttest), randomized	Unselected	Untreated	Study-developed program training WM of numbers, letters, colors, and spatial locations	15–45-min sessions over a period of 5 weeks
E. Dahlin, Nyberg, et al. (2008)	23.7 (23.4)	Pre-post randomized	Unselected	Untreated	Study-developed program training WM of numbers, letters, colors, and spatial locations	Training 3 times per week for 5 weeks, each session lasting 45 min
Holmes et al. (2009)	10.1 (9.9)	Pre-post Nonrandomized	Participants scored at or below the 15th percentile on two tests of verbal WM	Treated with computerized program	CogMed	35 min per day for at least 20 days over 5–7 weeks
Horowitz-Kraus & Breznitz (2009)	25.2 (25.2)	Pre-post follow-up (6 months after training) nonrandomized	Treatment group with dyslexia, normal reading controls		Cognifit	24 sessions, 15–20 min for 6 weeks
Jaeggi et al. (2008)	25.6 (25.6)	Pretest–posttest control group design, matched nonrandomized	Unselected	Untreated	Program developed for study, n-back training	Training from 8–19 days, daily training for about 25 min
Jaeggi et al. (2011)	9.1 (8.8)	Pretest–posttest control group, nonrandomized with follow up	Unselected	Treated with computerized language training	Adaptive spatial n-back training.	4 weeks, 5 times per week for 15 min
Jaeggi et al. (2010)						
Comparison 1	19.1 (19.4)	Pretest–posttest control group design, matched nonrandomized	Unselected	Untreated	Dual n-back training	Daily training 5 times per week for a period of 4 weeks
Comparison 2	19.0 (19.4)	Pretest–posttest control group design, matched nonrandomized	Unselected	Untreated	Single n-back training, only visuospatial tasks	Daily training 5 times per week for a period of 4 weeks
Klingberg et al. (2005)	9.9 (9.8)	Pre-post follow-up (3 months after training) randomized	ADHD (nonmedicated)	Treated with computerized program	CogMed	25 days of training, medium training time of 40 min for each trial, 5–6 weeks between pre- and posttest

(Appendix continues)

Table A1 (*continued*)

Study author & year	Age (years) E (C)	Design	Participant characteristics	Control treatment	Program type	Training duration
Klingberg et al. (2002)	11.0 (11.4)	Pre-post nonrandomized	ADHD (mix of medicated and nonmedicated)	Treated with computerized program (but for less time than the training group)	CogMed	25 days of training, medium training time 40 min for each trial, 5–6 weeks between pre- and posttest
Loosli et al. (2011)	10.0 (10.0)	Pretest–posttest control group design, matched nonrandomized	Unselected	Untreated	Visual working memory span task, n-back training	2 weeks of training, daily, 12 min per day
Nutley et al. (2011)	4.3 (4.3)	Pretest–posttest control group, randomized	Unselected	Treated	CogMed	5–7 weeks, 15 min per session for 25 sessions
Richmond et al. (2011)	66 (66)	Pretest–posttest control group, randomized	Older adults, participants with a minimal state exam score of 26 or below were excluded	Treated	Modeled after Chein & Morrison (2010)	20 30-min sessions over 4–5 days
Schmiedek et al. (2010) Comparison 1	25.6 (25.2)	Pre-post nonrandomized, matched on age, initial cognitive status and education	Unselected	Untreated	A mix of 12 tasks aimed at training working memory, memory, and processing speed	Mean number of training sessions: 101, mean time between pre- and posttest: 28 weeks
Comparison 2	71.3 (70.6)	Pre-post nonrandomized, matched on age, initial cognitive status and education	Unselected	Untreated	A mix of 12 tasks aimed at training working memory, memory, and processing speed	Mean number of training sessions: 101, mean time between pre- and posttest: 28 weeks
Shavelson et al. (2008)	13.5 (13.5)	Pre-post randomized	Unselected middle school children (8 with ADHD or learning difficulties)	Treated with computerized program	CogMed	5 days per week, 30–40 min per day, 25 days total
Shiran & Breznitz (2011) Comparison 1	25.1 (25.1)	Pretest–posttest control group, nonrandomized	Poor readers, university students with dyslexia	Self-paced reading intervention	Cognifit	24 sessions over 6 weeks about 15 min each
Comparison 2	24.8 (24.8)	Pretest–posttest control group, nonrandomized	Skilled readers, university students	Self-paced reading intervention	Cognifit	24 sessions over 6 weeks about 15 min each
St. Clair-Thompson et al. (2010)	6.10 (6.11)	Pre-post follow-up (5 months after training), nonrandomized	Unselected	Untreated	Memory Booster	2 × 30 min sessions per week for 6–8 weeks
Thorell et al. (2008) Comparison 1	4.5 (4.8)	Pre-post nonrandomized	Unselected	Treated with computerized program	CogMed	15 min daily for 5 weeks
Comparison 2	4.5 (5.0)	Pre-post nonrandomized	Unselected	Untreated	CogMed	15 min daily for 5 weeks
Van der Molen et al. (2010) Comparison 1	15.3 (15.4)	Pre-post follow-up (10 weeks after training), randomized	IQ in the range of 55–85	Treated	Computerized working memory training developed for study	6 min, 3 times per week over 5 weeks
Comparison 2	15.0 (15.4)	Pre-post follow-up (10 weeks after training), randomized	IQ in the range of 55–85	Treated	Computerized working memory training developed for study	6 min, 2 times per week over 5 weeks
Westerberg et al. (2007)	55.0 (53.6)	Pretest–posttest control group, randomized	Stroke patients	Untreated	CogMed	40 min a day, daily for 5 weeks

Note. C = control group; E = experimental training group; WM = working memory; ADHD = attention-deficit/hyperactivity disorder.

(Appendix continues)

Table A2

Outcome and Effect Size for Studies Included in the Meta-Analysis

Study author & year	Outcome construct (indicator)	Effect size <i>d</i> (pretest–posttest difference in gain)	Sample size pretest–posttest training (control)	Effect size <i>d</i> (pretest–follow-up test difference in gain)	Sample size pretest–follow-up training (control)
Alloway & Alloway (2009)	Verbal working memory (AWMA)	1.51**	8 (7)	—	—
	Verbal ability (Vocabulary WISC)	1.13*		—	—
	Arithmetic (Numerical Operations WOND)	0.58		—	—
Alloway (in press)					
Comparison 1	Verbal working memory (AWMA)	0.15	32 (39)	—	—
	Visuospatial working memory (shape recall test)	0.02		—	—
	Verbal ability (Vocabulary WASI)	–0.26		—	—
	Arithmetic (Numerical Operations WOND)	–0.11		—	—
Comparison 2	Verbal working memory (AWMA)	0.27	23 (39)	—	—
	Visuospatial working memory (Shape recall test)	0.66**		—	—
	Verbal ability (Vocabulary WASI)	0.49		—	—
	Arithmetic (Numerical Operations WOND)	0.17		—	—
Borella et al. (2011)	Verbal working memory (backward digit span)	2.09**	20 (20)	0.29	20 (20)
	Visuospatial working memory (dot matrix)	1.37**		0.11	
	Attention (Stroop)	0.67*		–0.08	
	Nonverbal ability (Cattell)	1.14**		0.79	
Chein & Morrison (2010)	Attention (Stroop)	0.56	22 (20)	—	—
	Nonverbal ability (Raven)	0.08		—	—
E. Dahlin, Nyberg, et al. (2008)					
Comparison 1	Verbal working memory (letter working memory task)	0.98*	15 (11)	1.01*	11 (7)
	Verbal ability (category fluency, e.g., say as many animals you can that start with the letter <i>s</i>)	–0.09		0.46	
	Nonverbal ability (Raven)	0.29		–0.14	
Comparison 2	Verbal working memory (letter working memory task)	1.12**	13 (16)	1.59**	13 (7)
	Verbal ability (category fluency, e.g., say as many animals you can that start with the letter <i>s</i>)	0.12		–0.08	
	Nonverbal ability (Raven)	0.06		0.28	
E. Dahlin, Nyberg, et al. (2008)	Verbal working memory (letter memory).	2.15**	15 (7)	—	—
	Attention (Stroop)	0.01		—	—
Holmes et al. (2009)	Verbal working memory (AWMA)	2.39**	22 (20)	—	—
	Visuospatial working memory (AWMA)	0.85**		—	—
	Nonverbal ability (WASI performance)	–0.19		—	—
	Verbal ability (Vocabulary WASI)	0.29		—	—
	Decoding (WORD basic reading)	–0.09		—	—
	Arithmetic (Numerical Operations WOND)	–0.11		—	—
Horowitz-Kraus & Breznitz (2009)	Verbal working memory (a number of adjectives in which the participants were to name their opposites in the same way as presented)	–0.16	27 (34)	–0.51*	27 (34)
	Decoding (1 min word and nonword decoding test)	0.00		0.09	
Jaeggi et al. (2008)	Verbal working memory (reading span task)	–0.07	26 (27)	—	—
	Nonverbal ability (BOMAT and Raven)	0.40	34 (35)	—	—
Jaeggi et al. (2011)	Nonverbal ability (Raven)	0.07	32 (32)	–0.04	32 (32)

(Appendix continues)

Table A2 (continued)

Study author & year	Outcome construct (indicator)	Effect size <i>d</i> (pretest–posttest difference in gain)	Sample size pretest–posttest training (control)	Effect size <i>d</i> (pretest–follow-up test difference in gain)	Sample size pretest–follow-up training (control)
Jaeggi et al. (2010)					
Comparison 1	Verbal working memory (N-back)	1.34**	25 (40)	—	—
	Nonverbal ability (BOMAT and Raven)	0.53	25 (43)	—	—
Comparison 2	Verbal working memory (N-back)	1.12**	20 (41)	—	—
	Nonverbal ability (BOMAT and Raven)	0.53	21 (43)	—	—
Klingberg et al. (2005)	Visuospatial working memory (the span board task)	0.77**	20 (24)	0.81*	18 (24)
	Nonverbal ability (Raven)	0.23	20 (24)	0.05	
	Attention (Stroop)	0.43	20 (23)	0.10	
Klingberg et al. (2002) ^a	Visuospatial working memory (the span board task)	1.66**	7 (7)	—	—
	Nonverbal ability (Raven)	2.18**		—	—
	Attention (Stroop)	1.28*		—	—
Loosli et al. (2011)	Nonverbal ability (TONI)	0.12	20 (20)	—	—
	Word decoding (Salzburger lesetest, words, and nonwords, accuracy)	0.34		—	—
Nutley et al. (2011)	Visuospatial working memory (the grid task)	1.55**	24 (25)	—	—
	Nonverbal ability (composite variable set A, AB, and B; Raven CPM; and block design WPPSI)	−0.17		—	—
Richmond et al. (2011)	Verbal working memory (Raven)	0.67*	21 (19)	—	—
	Nonverbal ability (Reading span)	−0.40		—	—
Schmiedek et al. (2010)	Verbal working memory (3-back numerical)	0.42*	101 (44)	—	—
Comparison 1	Visuospatial working memory (memory updating spatial)	0.12		—	—
	Verbal ability (verbal ability from Berlin Intelligence structure test)	0.13		—	—
Comparison 2	Nonverbal ability (Raven)	0.33		—	—
	Verbal working memory (3 back numerical)	0.10	103 (39)	—	—
	Visuospatial working memory (memory updating spatial)	0.11		—	—
	Verbal ability (verbal ability from Berlin Intelligence structure test)	−0.01		—	—
Shavelson et al. (2008)	Nonverbal ability (Raven)	0.54**		—	—
	Verbal working memory (composite operation span and reading span)	0.25	18 (19)	—	—
	Visuospatial working memory (span board task)	0.52		—	—
	Nonverbal ability (Raven)	0.01		—	—
Shiran & Breznitz (2011)					
Comparison 2	Visuospatial working memory (Cognifit test)	0.37	26 (15)	—	—
	Word decoding (words and nonwords decoding speed and accuracy)	0.36		—	—
Comparison 1	Visuospatial working memory (Cognifit test)	0.50	35 (15)	—	—
	Word decoding (words and nonwords decoding speed and accuracy)	0.40		—	—
St. Clair-Thompson et al. (2010)	Verbal working memory (listening recall working memory test battery for children)	1.28**	117 (137)	—	—
	Visuospatial working memory (block recall WMTB)	0.35*	69 (72)	—	—
	Arithmetic (WISC–IV Arithmetic)	0.25	46 (31)	0.27	44 (26)

(Appendix continues)

Table A2 (continued)

Study author & year	Outcome construct (indicator)	Effect size <i>d</i> (pretest–posttest difference in gain)	Sample size pretest–posttest training (control)	Effect size <i>d</i> (pretest–follow-up test difference in gain)	Sample size pretest–follow-up training (control)
Thorell et al. (2008) Comparison 1	Verbal working memory (forward and backward digit span)	1.09**	17 (14)	—	—
	Visuospatial working memory (the span board task)	0.45		—	—
	Nonverbal ability (block design WISC)	–0.03		—	—
	Attention (Day–Night Stroop task)	0.23			
	Comparison 2				
	Verbal working memory (forward and backward digit span)	1.06**	17 (16)	—	—
	Visuo-spatial working memory (the span board task)	0.70*		—	—
	Nonverbal ability (block design WISC)	0.33			
Van der Molen et al. (2010) Comparison 1	Attention (Day–Night Stroop task)	0.34		—	—
	Verbal working memory (composite backward digit recall, listening recall from WMTB)	0.16	41 (26)	0.13	39 (25)
	Visuospatial working memory (block recall WMTB)	0.17		0.42	
	Nonverbal ability (Raven)	–0.23		–0.23	
	Attention (Stroop)	0.20		0.10	
	Decoding (word decoding test)	0.09		0.17	
	Arithmetic (WISC–IV Arithmetic)	0.00		0.10	
	Comparison 2				
	Verbal working memory (composite backward digit recall, listening recall from WMTB)	0.20	26 (26)	0.06	25 (25)
	Visuospatial working memory (block recall WMTB)	–0.01		0.35	
	Nonverbal ability (Raven)	–0.23		–0.12	
	Attention (Stroop)	0.10		0.22	
	Decoding (word decoding test)	0.03		0.12	
	Arithmetic (WISC–Arithmetic)	0.05		0.17	
	Westerberg et al. (2007)				
	Visuospatial working memory (span board Wechsler)	0.78	9 (9)	—	—
	Nonverbal ability (Raven)	–0.10		—	—
	Attention (Stroop)	–0.29		—	—

Note. AWMA = Automated working memory assessment; WASI = Wechsler Abbreviated Scale of Intelligence; WISC = Wechsler Intelligence Scale for Children; WISC–IV = Wechsler Intelligence Scale for Children—Fourth Edition; WOND = Wechsler Objective Numerical Dimensions; WORD = Wechsler Objective Reading Dimensions; WMTB = Working Memory Test Battery for Children; BOMAT = Bochumer Matrices Test; TONI = test of nonverbal IQ; CPM = colored progressive matrices; WPPSI = Wechsler Preschool and Primary Scale of Intelligence.

^a Standard deviations used to calculate the effect sizes are estimated on the basis of standard errors for the means reported in the article.

* $p < .05$. ** $p < .01$.

Received September 16, 2011

Revision received February 21, 2012

Accepted February 27, 2012 ■